

Rubric for Assessment of Learning Process in EFL Online Classrooms: Validation and Revision¹

Faisal Mustafa², & Siti Raisha³, Universitas Syiah Kuala, Banda Aceh, Indonesia
Mahlil⁴, Politeknik Negeri Lhokseumawe, Lhokseumawe, Indonesia

Abstract

Assessment of the learning process refers to assessing the quality of students' learning as they complete learning activities, such as how much time they spend reading materials, how many times they repeat quizzes when they get low scores, or whether their posts in a forum are helpful for other students. Assessment of process is more appropriate for online learning, and the use of a rubric has been suggested to ensure the objectivity of learning process assessment (Wilén, 2004). The objective of the present study was to validate such a rubric for the assessment, proposed by Mustafa et al. (2019) on the basis of feedback solicited from students. Sixty percent of the students who had successfully completed and passed a compulsory online class, English as a foreign language (EFL), were recruited to participate in the study ($N = 72$). The data consisted of a questionnaire which explored whether the rubric grades (for each criterion and the overall passing grade) were adequate from the perspective of students. It also explored whether or not the rubric grade they received matched students' expectations; significant differences between the rubric and students' expectations were determined quantitatively. Both grades were compared using the one-sample Wilcoxon sign rank test at the significance level of 0.01. The results showed that the grades for seven out of twenty criteria in the rubric required revision because they were significantly different from the grades proposed by the students. The new revised version of the rubric can be used to improve student satisfaction with their assessment results in an online learning classroom.

Resumen

La evaluación del proceso de aprendizaje se refiere a evaluar la calidad del aprendizaje de los estudiantes a medida que completan las actividades de aprendizaje. Incluye cuánto tiempo dedican a leer materiales, cuántas veces repiten las pruebas cuando obtienen puntajes bajos o si sus publicaciones en un foro son útiles para otros estudiantes. La evaluación del proceso es más apropiada para el aprendizaje en línea, y se ha sugerido el uso de una rúbrica para asegurar la objetividad de la evaluación del proceso de aprendizaje (Wilén, 2004). El objetivo del presente estudio fue validar dicha rúbrica para la evaluación, propuesta por Mustafa et al. (2019) sobre la base de los comentarios solicitados a los estudiantes. El sesenta por ciento de los estudiantes que habían completado y aprobado con éxito una clase obligatoria en línea, inglés como lengua extranjera (EFL), fueron reclutados para participar en el estudio ($N = 72$). Los datos consistieron en un cuestionario que exploró si las calificaciones de la rúbrica (para cada criterio y la calificación general aprobatoria) eran adecuadas desde la perspectiva de los estudiantes. También exploró si la calificación que recibieron coincidió con las expectativas de los estudiantes; Las diferencias significativas entre la rúbrica y las expectativas de los estudiantes se determinaron cuantitativamente. Ambos grados se compararon utilizando la prueba de rango de signos de Wilcoxon de una muestra al nivel de significancia de 0.01. Los resultados mostraron que las calificaciones para siete de veinte criterios en la rúbrica requerían revisión porque eran significativamente diferentes de las calificaciones propuestas por los estudiantes. La nueva versión revisada de la rúbrica puede utilizarse para mejorar la satisfacción de los estudiantes con los resultados de sus evaluaciones en un aula de aprendizaje en línea.

Introduction

Electronic learning, or e-learning, has been used in all fields of study, including English language teaching (Bataineh & Mayyas, 2017). Research has revealed many benefits of e-learning in English teaching (Bataineh & Mayyas, 2017; Pop & Slev, 2012; Shams, 2013). For example, EFL learners who do not have access to authentic materials can use videos available on the internet. Research has also revealed that students become more autonomous learners through e-learning (Shams, 2013). Furthermore, e-learning allows teachers to bring native speakers into their EFL/ESL classrooms virtually.

Most significantly, e-learning also enables immediate assessment (Fageeh, 2015). Many language-test designers have developed e-raters for evaluating productive, objective skills such as writing and speaking (Attali & Burstein, 2004). E-raters can be added to e-learning platforms, such as Moodle, for direct classroom use (Bateson, 2019). With immediate feedback, students are able to identify their weaknesses and autonomously improve them (Fu & Li, 2020). Although this type of assessment has many weaknesses, such

¹ This is a refereed article. Received: 21 May, 2020. Accepted: 13 March, 2021.

² faisal.mustafa@unsyiah.ac.id

³ siti.raisha@unsyiah.ac.id

⁴ mahlil@pnl.ac.id

as providing only partial coverage of student development and lack of transparency (Quinlan et al., 2009), it is preferable to traditional teacher-led assessment, where such immediate feedback is not possible at all.

Unlike assessment for testing purposes, such as quizzes or exams, teachers also need to evaluate students' learning process. Kovalik and Dalton (1999) defined such practices as the assessment of process, that is, the assessment of how students complete learning activities by examining the quality of their learning. Assessment of process should cover all types of learning activities, which include reading assigned material, completing exercises, and participating in group or class discussions. Mustafa et al. (2019) proposed that the grade given for learning process assessment should be based on the effort that students put forth in completing a learning activity. They provide a rubric for assessing the process of reading materials and completing exercises. The rubric considers time spent, number of attempts, and grades obtained. However, the rubric has not been validated. Therefore, the current study validates this rubric by determining whether students agreed with the initial rubric grades. When they disagreed, students were asked to propose the grade they considered appropriate for each assessed criterion. Therefore, this study analyzes whether the grades proposed by the students differ significantly from those provided in the existing rubric. When the grades were significantly different, they were revised based on the students' feedback for the proposed grade. Only the grade for each criterion was revised, while the criteria themselves were retained. The results of this study enable the researchers to revise and design a more complete assessment rubric for online learning that also accounts for the perspective of students. The revised rubric allows EFL and other foreign-language teachers to objectively assess their students' learning process. Teachers can also distribute the rubric to students to better inform them of learning process expectations; thus students can put forth their best effort to meet those expectations.

The paper proceeds with a literature review related to the variables in this study. The second section describes the methodology employed for this study, followed by a presentation and discussion of the results.

Literature review

Assessment of process

Assessment is the final phase in the teaching process, determining future improvements for the teaching process and teacher development (Griffin & Care, 2015). Therefore, Griffin (2017) suggested that assessment must be as reliable and accurate as possible to produce information that is helpful for both teachers and students. Brown (2004) categorized assessment into two broad types, that is, summative and formative assessment. Formative assessment is conducted during the learning process to help students learn more effectively, while summative assessment measures how well students have learned (Yulia et al., 2019).

Assessment of the learning process, a part of formative assessment, provides teachers and students with input on the learning process and whether the desired outcomes were met (Elliott & Yu, 2013). Through the assessment process, teachers collect initial information about the needs and problems encountered by students in the learning process. As a result, teachers can alter, adjust, and incorporate further learning material, if necessary (Moss & Brookhart, 2009). Therefore, learning process assessment evaluates the shortcomings or problems learners experience. It helps a teacher consider best approaches for tackling such problems, not only for students' achievement but also for their learning experience. Teachers normally conduct learning assessment, but research has proven that self-assessment by students can also improve their learning performance (Earl & Katz, 2006; Sharma et al., 2016). Students may also be evaluated by peers who provide input on the quality of their academic development (Double et al., 2019). The results of peer assessment help students improve their learning habits and allow teachers to adjust their instructional approach (Khonbi & Sadeghi, 2013).

With the use of technology, teachers require an approach to assessment that differs from its traditional form (Alverson et al., 2019). Online assessment records the time spent on reading material and completing exercises, as well as the number of attempts and the grade obtained for each attempt (McCracken et al., 2012). In addition, this method documents students' participation in online discussion forums, and teachers can immediately assess how students read the materials (Arkorful & Abaidoo, 2015; Kayler & Weller, 2007; Ozcan-Seniz, 2017). Formative and summative assessments can identify learning processes and outcomes through e-learning platforms such as Blackboard, Canvas, or Moodle (Gamage et al., 2019).

Process assessment for lesson

Process assessment for lessons commonly involves evaluating the students' learning processes as they read the materials provided by a teacher (Sieberer-Nagler, 2015). It includes observing how well the students understand the learning material or lessons provided. Teachers must ensure that their students do not copy answers from their peers; they must read the material before answering the questions provided in classroom activities (Tauber, 2007). In classroom management practice, teachers are encouraged to make sure students have read the material by walking around the room and checking on each student briefly (Tauber, 2007). To encourage students to read the materials before class, teachers can provide reading questions as a homework assignment (Hoeft, 2012).

Several studies have stated the importance of assessment for the learning process (Almazán, 2018; Espinosa Cevallos & Soto, 2018; Rodrigues & Oliveira, 2014). Feedback first requires assessment. Earl and Katz (2006) reported that feedback provided in an assessment helps students to improve their learning because the assessment is conducted during the learning process. This opportunity helped low-achieving students to improve their understanding of materials. Dennen (2005) stressed that feedback given during the learning process better motivates students.

Assessments for exercises

An exercise or a quiz is an important component of the teaching and learning process (Mustafa, 2015). Exercises are even more important for EFL students because language learning is more skills-based than knowledge-based (Richards & Rodgers, 1986). Exercises in the classroom help students learn and recall the learning material (Brown, 2004); they are in themselves a type of formative assessment. Feedback for such assessment can be obtained in scores or comments, depending on the nature of the exercise (Falchikov, 2007). However, a more comprehensive assessment is possible for exercises delivered through e-learning. One of the advantages of the online practice test is immediate feedback and the possibility to reattempt the exercise based on feedback (Padayachee et al., 2018). Research on the number of ideal attempts showed that most students attempt exercises two or three times. The number of correct answers increases with each attempt, while the completion time decreases (Cohen & Sasson, 2016). According to Gamage et al. (2019), a good quiz intended for formative assessment should give students an opportunity to practice and apply the materials they learn.

Assessment for participation in discussion

The learning process also includes discussion among students and teachers, and students with peers. In a regular classroom, Borich (2013) described that discussion occupied about 80 percent of all school time. Studies report that students with better participation in classroom discussions obtain better scores in the final test compared to those with lower participation (Kim et al., 2020). A similar effect was observed in an online classroom (Davies & Graff, 2005) where the discussion provided an opportunity for students to "make sense" of what had been read (Nor et al., 2010). Therefore, teachers must help students participate in classroom discussions. Another study revealed that assessment encourages students to participate more intensely in discussions (Balaji & Chakrabarti, 2010). Further, to assess participation in discussion forums, Wang and Chen (2017) stated that the instructor must align the discussion, course objective(s), and assessment. This more effectively communicates to the students the purpose of the discussion, weaves the discussion into the assignment, and builds interdependence into the assignment. They further suggest that teachers should grade participation based on the quality of students' posts. Wilen (2004) proposed the use of a rubric for a more objective assessment of discussion.

Rubric for the assessment of learning process in online classrooms

For reliability and objectivity, the learning process must be assessed using a rubric (Al-Rabai, 2014; Robles et al., 2014). A study conducted by Mustafa et al. (2019) revealed that vocabulary homework delivered using e-learning differed from paper-and-pencil vocabulary homework. In the treatment (i.e., e-learning) group, the students were required to read materials consisting of definitions and explanations of all vocabulary. They were then required to complete all the exercises provided. The students' effort in reading the materials and completing the exercises was graded using a scoring rubric, as presented in Table 1 and Table 2 below.

No	The status on e-learning	Description	Grade
1	No status	Student has not clicked the lesson.	0
2	Lesson not started	The student has clicked the lesson.	1
3	Course module viewed/ Content page viewed	The student has started to view the lesson but has not finished reading all the content in the lesson.	2
4	Lesson ended	The student has finished reading the lesson.	3

Table 1: Rubric for the assessment of learning process of lessons (Mustafa et al., 2019)

No	The status on e-learning	Description	Grade
1	No status	Student has not clicked the exercise	0
2	Course module viewed	The student has clicked the exercise.	0
3	Quiz attempt started	The student has started the exercise.	1
4	Quiz attempt viewed	The student has finished the exercises but has not clicked on the "submit" button.	2
5	Quiz attempt summary viewed	The student has reviewed the exercise before submission.	2
6	Quiz attempt submitted	The exercise was submitted.	3
7	The exercise has been submitted, but the completion duration is too short.		1.50
8	The submitted quiz is less than 80% correct, but the student did not reattempt the exercise.		1.75

Table 2: Rubric for the assessment of learning process of quizzes/exercises (Mustafa et al., 2019)

The rubric designed by Mustafa et al. (2019) met the criteria proposed by Wolf and Stevens (2007). It consisted of criteria with a minimum of three levels and performance descriptions. Mustafa et al. (2019) originally used the rubric to assess the students' learning process in completing vocabulary homework. They compared the scores of students who completed the vocabulary homework to those who did not. They then concluded that students who completed the vocabulary homework outperformed those who did not complete the homework in a reading comprehension test. However, to reflect students' real learning activities, a rubric must be validated (Raposo-Rivas & Gallego-Arrufat, 2016). When a rubric is intended to assess the learning process of students, the rubric developers should consider students' feedback (Andrade, 2000). Therefore, the rubric designed by Mustafa et al. (2019) needs to be validated based on students' feedback and revised accordingly. Consequently, the present study addressed whether the grades proposed by the students significantly differed from those provided in the existing rubric.

Methodology

This study drew on mixed methods to develop a revised rubric for the assessment of e-learning, based on the original rubric devised by Mustafa et al. (2019). Quantitative methods were used to test for the statistical significance of the differences between the grades proposed by the students and those provided in the rubric. Based on the results of this quantitative analysis, a qualitative analysis was conducted and the rubric was revised. First, quantitative results were analyzed using inferential statistics, in this case a one-sample Wilcoxon sign rank test. Then, data were compared with qualitative data, that is, students' perceptions of the grade proposed in the rubric for each learning activity. Qualitative data was collected by means of a questionnaire.

Participants

The questionnaire was administered to five groups of students at Universitas Syiah Kuala, Indonesia. Most of the participants had a low level of English proficiency, but they had formally learned English for approximately 750 hours, on average, throughout high school and university. The course was an extra graduation requirement for students who could not obtain the minimum TOEFL (Test of English as a Foreign Language) score of 477. Successful completion of the course was determined using the modified rubric previously proposed by Mustafa et al. (2019). Thus, the rubric was used as a formal assessment. It was the only assessment which was used to determine whether or not the students passed the course. Students who did not successfully complete the course were required to repeat it. Only students who passed the course without any repetition were recruited as participants in this study. Thus, the total number of participants was 72.

Data collection

The questionnaire administered in class consisted of three sections: lesson, quiz, and forum. The questionnaire was based on the rubric used in class (see Appendix 1). In the questionnaire, the participants were asked whether the grade given for each learning activity was adequate, high, too high, low, or too low. After that, the participants were requested to provide the grade they considered most appropriate for each learning activity, which was between 0 and 3 in 0.25-point increments.

The questionnaire was administered anonymously. The participants were enrolled in the TOEFL training as a requirement for their undergraduate degrees. Instruction in listening comprehension, structure, and reading comprehension was delivered through face-to-face interaction by different instructors. However, due to COVID-19, the classes moved online. Online classes took place on the platform Moodle 3.0, which was accessible via the network which was accessible via internet provided by the university. Students were familiar with the platform because the university used a similar platform for their curricular classes. After the students completed the course, they were asked to complete the questionnaire, and they were told their responses were anonymous. The return rate was over 95%. All participants who completed the questionnaire gave their consent to use the results of the questionnaire for research purposes.

Data analysis

To decide whether the existing grades needed to be revised, the grades (for each criterion in the rubric and the overall passing grade) were compared to the grades proposed by students on the questionnaire using the one-sample Wilcoxon sign rank test (Privitera, 2018). For accuracy and time-efficiency, all statistical analyses were performed using R, an advanced open-source statistical programming application. The significance level used in the analysis was 0.01. The low alpha level was used to avoid the risk of a type I error (Heumann et al., 2016). If the p -value for any grade was lower than 0.01, meaning that the existing rubric grade was significantly different from the student-proposed grade, it was revised using the procedure presented as follows.

Revision process

The expectation was that the rubric grades for the overall passing score and for each criterion would not be significantly different from the grades proposed by the students. If the one-sample Wilcoxon sign rank test revealed a significant result ($p < 0.01$), then the grade in the rubric was increased or reduced by 0.25. If the p -value approached 0.01, the grade was increased or reduced by another 0.25 until the p -value reached 0.01. After adding or reducing the grade by 0.25, the one-sample Wilcoxon sign rank test was repeated. If the p -value did not reach 0.01 after repeated calculation, the existing grade in the rubric was used in the revised rubric.

Results and Discussion

The objective of the study was to determine students' perceptions of the grades provided for each assessed criterion of the e-learning process. It also analyzed whether the grades proposed by students differed significantly from the grades in the existing rubric. Before the student questionnaire was administered, students were assessed using the original rubric proposed by Mustafa et al., (2019). The results of this assessment were used to determine whether the students passed or failed the training. The results were also announced before the questionnaire was distributed. Therefore, when completing the questionnaire, students had a clear picture of the rubric's importance. The information provided by the students was used to revise the rubric. The data were analyzed for each type of learning activity, and the results are presented separately.

Grades for a lesson activity

The first questionnaire section sought students' perceptions of the grade given for each learning activity. Activities were categorized as lessons, quizzes or exercises, and forum participation. Lessons consisted of materials which students read in order to complete related exercises. Students' perceptions of grades for the lesson are provided in Figure 1 below, in which TL stands for "too low", L "low", A "adequate", H "high",

and TH "too high." For example, "too low" means that students thought the score for the assessed criterion underestimated their actual efforts and they believed the rubric score should be increased.

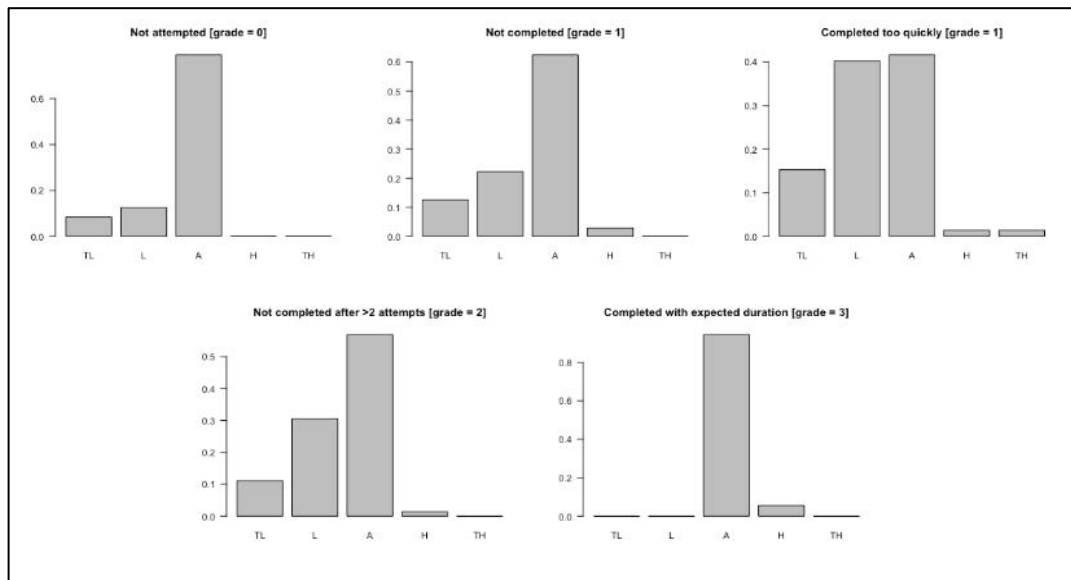


Figure 1: Students' perception of lesson grades

Figure 1 shows that most of the students agreed with the grade provided in the rubric, except for grades that were lowered for completing a lesson too quickly. Several students reported the score was either low or too low. In fact, many students in this experiment completed reading lessons too quickly. Previous studies also report that, when asked whether e-learning reading materials were easy to read and understand, more than half of student participants did not completely agree (Coiro & Dobler, 2007; Leyva, 2003). In addition, students easily develop a negative attitude toward reading (Chapman & Tunmer, 2003; Ganske et al., 2003). According to Margolis and McCabe (2004), this negative attitude could result in a failure to read the materials required for tests or homework. In this case, Indonesian students' reading habits were unsatisfactory, according to the Progress in International Reading Literacy Study (PIRLS) test (Tri & Ratri, 2015). The following table shows the difference between the provided grades and the grades proposed by the students for each activity related to lesson completion.

Rubric	given	proposed	p-value	revised	p-value
Not attempted	0	0.212	< 0.001	0.25	0.016
Not completed	1	1.225	< 0.001	1.25*	0.292
Completed too quickly	1	1.299	< 0.001	1.25*	0.550
Not completed after >2 attempts	2	2.125	0.016	-	-
Completed with exp. Duration	3	2.934	0.004	2.75	< 0.001

Table 3: Proposed lesson grades

As Table 3 shows, some students expected to receive a grade even though the lesson had not started. However, the percentage of such students numbered less than 15 percent. The grade that they proposed was not incorporated into the revised rubric because they did not deserve any credits for ignoring the lesson. This is expected because previous research finds that students normally read the material if they are given a follow-up quiz based on the assigned material (Cook & Babon, 2017). Second, a small number of students considered that 3 was a high grade for completing the material within the expected time limit; however, lowering the grade to 2.75, as suggested by students, made its comparison with the rubric grade significantly different ($p < 0.01$). Therefore, the rubric retained a grade of "3" for completing material within the expected time limit. Only the grades for an incomplete lesson or too-quick completion were revised from 1 to 1.25, making them similar to those proposed by the students. For the item *not completing the lesson after more than three attempts*, the analysis shows there was no evidence that the student-proposed grade and the rubric grade were significantly different ($p > 0.01$). Thus, the rubric grade was not revised. The following figure shows a summary of grades proposed for each item in the rubric.

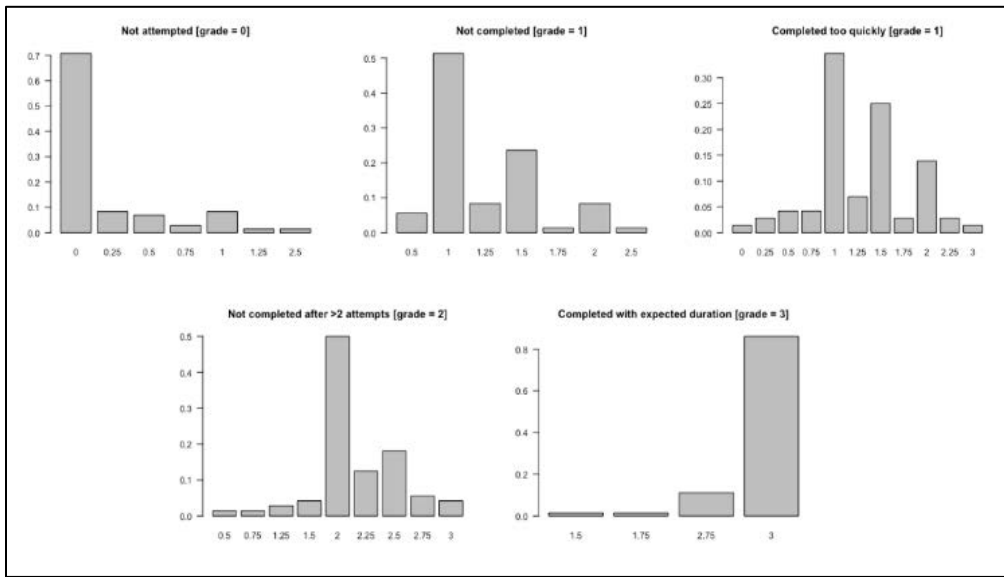


Figure 2: Students' suggested lesson grades

Although students considered the grades for some items as low or too low, most of them agreed with the provided grades. Figure 2 shows that a wide range of grades were suggested for both uncompleted lessons and lessons completed too quickly, which confirms that the grades for both items needed to be revised.

Grades for quiz and assignment

Quizzes and assignments assigned in the course were not used for testing purposes, rather, they were treated as learning activities, as described by Cohen and Sasson (2016). Therefore, both the duration and score were considered in determining the grade for these types of learning activities. Figure 3 presents the students' perception of each type of activity completion.

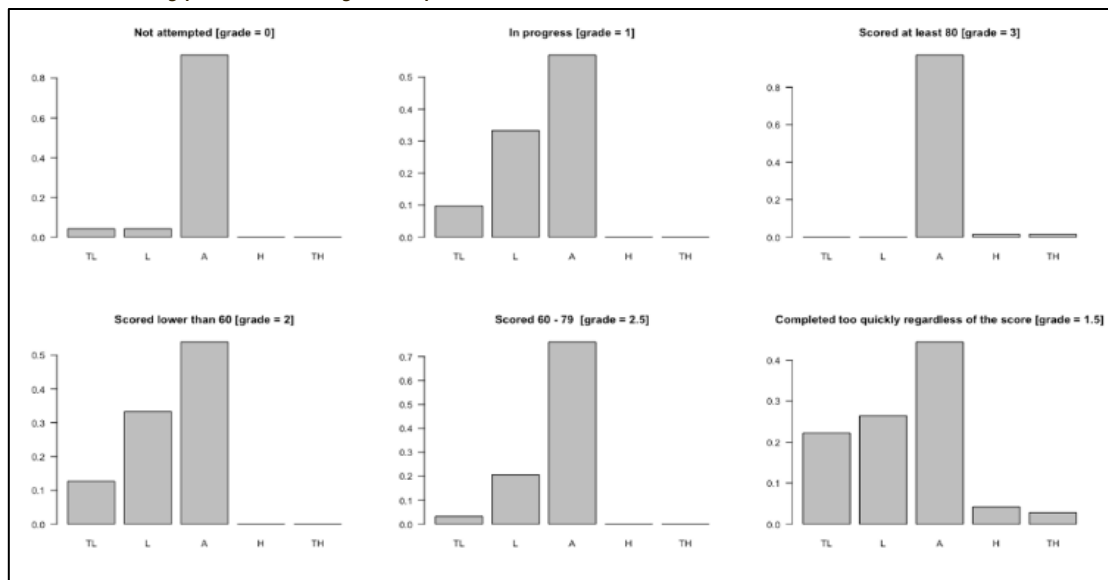


Figure 3: Research results for students' perception of quiz grades

Figure 3 shows that many students considered half of the grades for quizzes to be low or too low. While students agreed that no grade should be given if a quiz has not been opened, many of them reported that unsubmitted quizzes should have been graded higher than what the rubric provided, as should the other two items. Therefore, some revisions were made, as shown in Table 4 below.

Rubric	Given	proposed	p-value	revised	p-value
Not attempted	0	0.069	0.022	-	-
In progress	1	1.212	< 0.001	1.25*	0.142
Scored at least 80	3	2.965	0.048	-	-
Scored lower than 60	2	2.234	< 0.001	2.25*	0.699
Scored 60 – 79	2.5	2.556	0.011	-	-
Completed too quickly	1.5	1.747	0.000	1.75*	0.744

Table 4: Proposed quiz and assignment grades

The results presented in Table 4 are in line with data in Figure 3, that is, three items needed to be revised. The item “completing the quiz too quickly” was graded lower in the rubric because it indicates that cheating has taken place, as described in Galizzi (2010). However, 48.61% of the students considered the score as low or too low. In addition, Cohen and Sasson (2016) found that the time spent completing quizzes online was not a predictor of students’ performance on a proctored exam. This indicates that our initial prediction was not as accurate, and thus we revised the grade for this item. All revisions involved adding 0.25 to the rubric grades. The grades that students proposed for the other three items were statistically similar to those provided in the rubric ($p > 0.01$); thus no revision was necessary. Students agreed that the highest grade should not be given for quiz scores less than 80. The summary of grades proposed by students is provided in Figure 4.

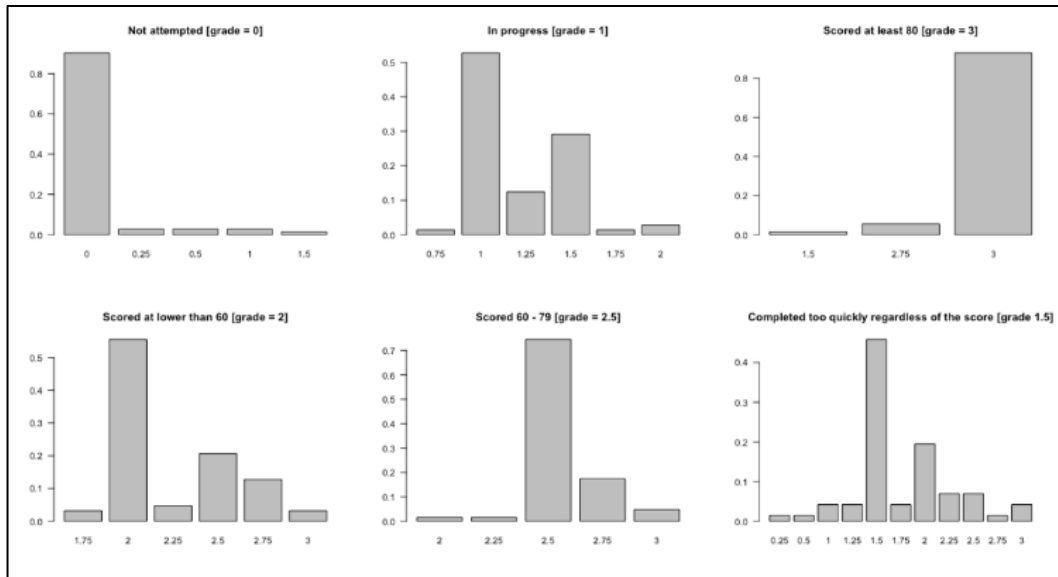


Figure 4: Students’ suggested quiz grades

Figure 4 demonstrates that three items in the rubric needed to be revised. Many students proposed higher grades for unsubmitted quizzes, quizzes scoring lower than 60, and quizzes completed too quickly. This result aligns with the students’ perception of quiz grades as presented in Figure 3. It is also supported by the statistical analyses given in Table 4.

Grades for forum participation

Students were expected to participate in an e-learning forum in lieu of participating in a real classroom discussion. Participation can include addressing classmates, asking questions, answering questions, and confirming information or materials. The rubric provided has taken these forms of participation into account and assigned a grade for each form of participation. The students gave their opinion for the grades in the rubric, whether they considered the grades too low (TL), low (L), adequate (A), high (H), or too high (TH). Figure 5 presents the percentage of students who gave their opinion for the grade of each form of participation.

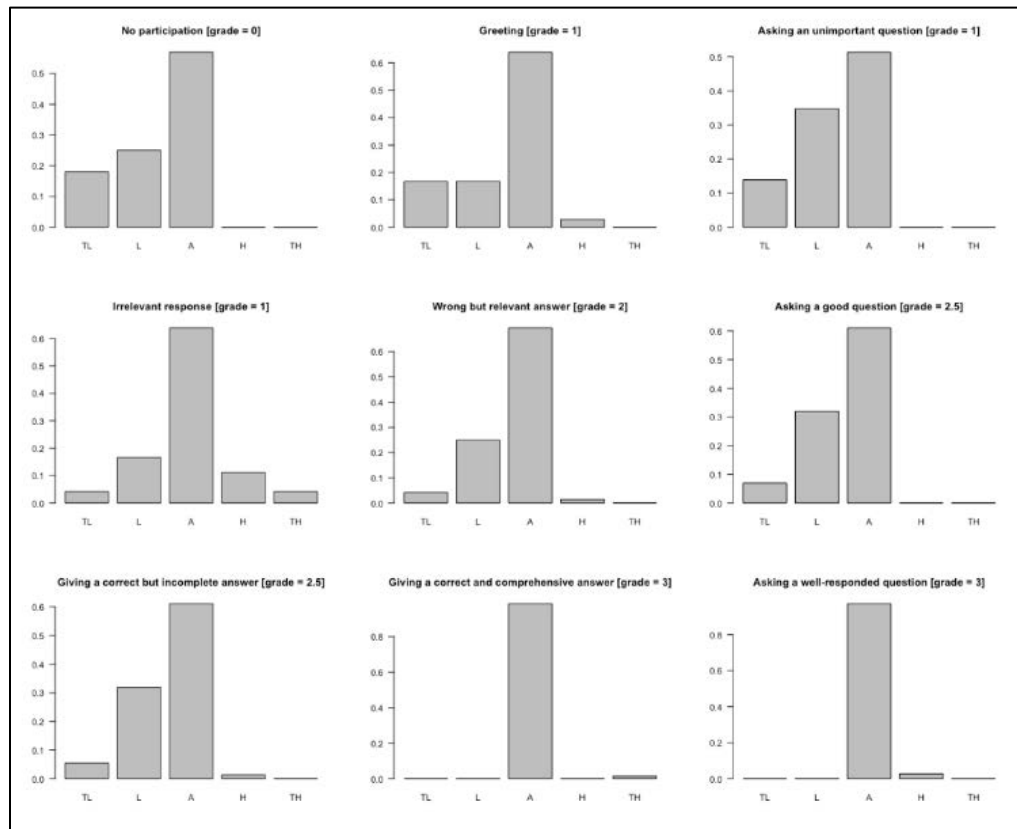


Figure 5: Students' perception of forum participation grades

The figure above shows that many students had conflicting views on at least five items in the rubric grade for forum participation. Two items with a grade of "3" received positive responses. For giving an irrelevant answer, the grade of "1" was perceived as low or too low by almost as many as those who perceived it as high or too high. Revisions for forum participation grades are presented in Table 5.

Rubric	given	proposed	p-value	revised	p-value
No participation	0	0.385	< 0.001	0.25	< 0.001
Participation by greeting	1	1.191	< 0.001	1.25*	0.054
Asking unnecessary questions	1	1.420	< 0.001	1.50*	0.212
Giving an irrelevant response	1	1.045	0.359	-	-
Giving wrong but relevant answers	2	2.146	< 0.001	2.25	< 0.001
Asking a good question	2.5	2.649	< 0.001	2.75	< 0.001
Giving a correct but incomplete answer	2.5	2.618	< 0.001	2.75	< 0.001
Giving a correct and complete answer	3	2.993	0.346	-	-
Asking a well-responded question	3	2.969	0.097	-	-

Table 5: Proposed grades for forum participation

Students' views on appropriate participation grades differed from the rubric. However, the revised grades students proposed did not significantly differ from the rubric. The data can be summarized in three points. First, the proposed grades of three items were statistically similar to the grades in the rubric. Thus, no revision was required. Secondly, four items were significantly different in the rubric versus student-proposed grades. However, revising the grades did not change the results ($p < 0.01$); thus the grades in the rubric were maintained. Finally, the grades for the other two items, that is, *greeting* and *asking a question which has already been answered*, were significantly different from grades proposed by the students. Therefore, the grades were revised by adding 0.25 for one item and 0.5 for the other, which made the grades statistically similar to what the students proposed ($p > 0.01$). That a minimum number of items needed to be revised illustrates that students believed the quality of discussion forum posts is essential. In fact, Wang

and Che (2017) found the quality of posts impacted students' learning achievement. The complete summary of forum participation grades proposed by the students is provided in Figure 6 below.

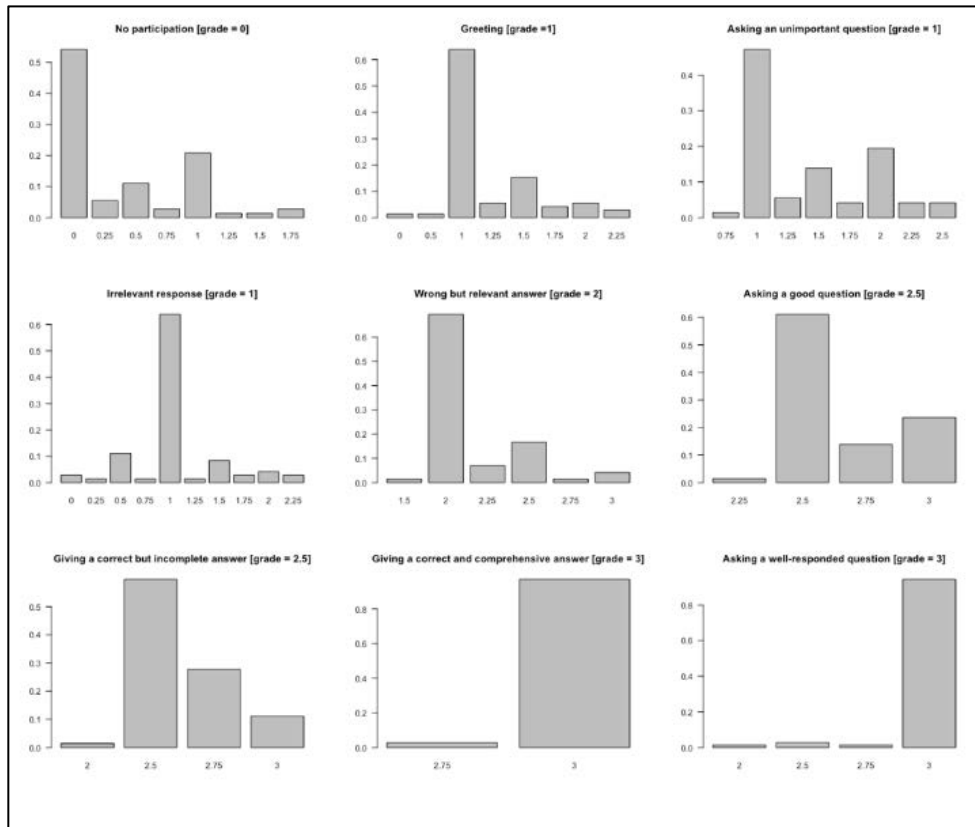


Figure 6: Students' suggested forum participation grades

Figure 6 shows a wide range of disagreements among students about the grades for the second and third items in the rubric. This highlights that revision was necessary for these two items.

The passing grade

The cumulative grade was calculated by taking the average of all grades in the course. In this study, we used 2.4, as described in the previous section, as the passing grade. The students' perception of this passing grade, and their proposed passing grades are presented in Figure 7.

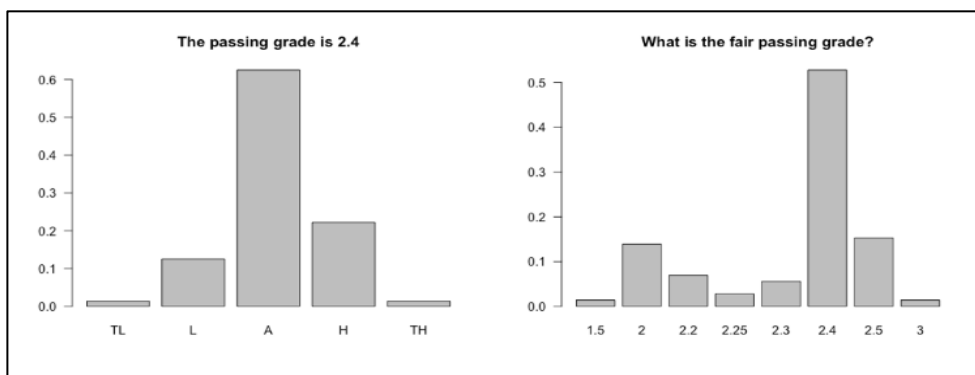


Figure 7: Assessment of passing grades

While many students perceived 2.4 as low or high, 60% of the students considered it an adequate passing grade. When asked to provide a passing grade, 30% of the students proposed grades below 2.4, while only 17% preferred the grade above it. Meanwhile, 53% of students agreed with the existing grade of 2.4. The mean of student-proposed grades was 2.332, which is significantly different from 2.4 ($p < 0.01$). Therefore, the passing grade was lowered to 2.35, which remains 2.4 when rounded. This grade was not significantly

different from the students' proposed grade, with a p -value of 0.668. When this score was converted to a percentage using cross product calculation, 2.35 was equivalent to 78%, while 2.4 was equivalent to 80%. The new passing grade is categorized as "distinction" in the Indonesian higher education system (Siagian & Siregar, 2018). In China, the grade is categorized as average, while in Canadian universities, the grade is considered "very good" (Hohner & Tsigaris, 2010). Furthermore, in Nigeria the grade of 78 falls into the category of "2nd Class Upper", which is one point lower than high distinction (Omotosho, 2013). Therefore, the passing grade of 78 or 2.35, as proposed by the students, is an adequate passing grade for overall assessment of process.

Overview of the results

The results of this study provide some significant contributions regarding applications of e-learning in language classrooms. First, many of the e-learning students in this study did not like reading the materials provided by their teacher. This conclusion is based on students' perceptions of the grade given for students who did not open the quiz. They proposed that students should receive a grade between 0.25 and 2.5 for a lesson, although they had not even started it. In addition, the students expected good grades for a lesson they had not finished or started without reading. Therefore, it is suggested that the e-learning materials be made as brief as possible and delivered through video. The teachers can record themselves presenting the materials, or an online video conference can be held to teach or discuss the materials with the students.

Students also expected high grades even when they finished a quiz too quickly. Throughout this e-learning course, we found that many students completed quizzes too quickly. For instance, they spent 10 minutes or less to complete 30 reading comprehension questions, yet they attained very high scores. The teachers suspected this approach indicates they had already received answers from peers, and they input the answers without reading the texts (Mellar et al., 2018). In addition, students might have translated English texts into their first language using a translation tool such as Google Translate. Therefore, one solution could be for teachers to shuffle questions and answer choices to inhibit cheating, and reading texts could be replaced with images so materials cannot be copied for translation or translated automatically in an internet browser.

Another finding is that many students did not appear to enjoy participating in the forum, but they considered the grade to be low or too low for non-participating students. This lack of participation is similar to many non-online classes (Hamouda, 2012; Khoiasteb et al., 2015; Tesfaye Abebe & Deneke, 2015). In an online class, participation is recorded and can be graded according to the quality of participation. In this study, we also found that students expected higher grades for low-quality participation, such as asking unnecessary questions. According to Juniati et al. (2018), a lack of participation is a result of a lack of confidence, a lack of motivation, emotional problems, and a lack of understanding. Cultural factors can also affect class participation (Savaşçı, 2014).

In such circumstances, teachers should give students difficult assignments for more compelling discussion. In addition, group work can be assigned so the students are forced to discuss online. This is easily facilitated in e-learning platforms such as Moodle. Furthermore, teachers can use an opening question to entice students to respond and participate in discussions. There are three types of opening questions which are effective in an online environment: exploration (a question that allows open-ended exploration of real phenomena), concept invention (a question on a new concept such as the definition of new vocabulary), and application (a question presenting problems to be solved using a new concept). According to Dalelio (2013), the application question is the most suitable type of opening question in a language learning classroom because it attracts more posts in an online discussion.

Students were satisfied with a grade of "zero" for not attempting a quiz. This suggests that the students considered completing a quiz to be more important than reading a lesson or participating in a discussion forum. In fact, based on a survey conducted by Salas-Morera et al. (2012), students spent twice as much time preparing and completing quizzes as participating in a discussion forum, and students also considered quizzes the most helpful for their learning. Another expected result is that students agreed that good-quality online posts deserve high grades. They also attract responses from other students (Swan et al., 2019).

Conclusion

In summary, the study validated the modified rubric for online learning assessment initially proposed by Mustafa et al. (2019). Based on the quantitative analysis, the grades of three items in the lesson assessment were statistically different from those proposed by the students. Meanwhile, two grades in the rubric for

quiz assessment and two for discussion forum participation were significantly different from the students' suggestions. These results led to a revision of two items in lesson assessment, three items in quiz assessment, and two items discussion forum assessment. The revised version of the rubric is attached in Appendix 1.

Limitations of the study

This study has revealed that students believed grades given for quizzes submitted too quickly should have been higher than what was given in the rubric. A lower grade was initially given because such a short duration of completion often indicates cheating, such as copying answers from peers. In the current study, we did not collect any qualitative data to more accurately predict how quizzes were completed, and whether cheating was involved. Therefore, we decided to increase the grades following students' suggestions. Further research should interview students to determine the real reasons some students completed quizzes too quickly.

References

- Al-Rabai, A. A. (2014). Rubrics revisited. *International Journal of Education and Research Rubrics*, 2(5), 473–484. <https://www.ijern.com/journal/May-2014/39.pdf>
- Almazán, J. T. (2018). An interview on assessment with Randall Davis. *MEXTESOL Journal*, 42(2), 1–3. http://www.mextesol.net/journal/index.php?page=journal&id_article=3492
- Alverson, J., Schwartz, J., & Shultz, S. (2019). Authentic assessment of student learning in an online class: Implications for embedded practice. *College and Research Libraries*, 80(1), 32–43. <https://doi.org/10.5860/crl.80.1.32>
- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13–18.
- Arkorful, V., & Abaidoo, N. (2015). The role of e-learning, advantages and disadvantages of its adoption in higher education. *International Journal of Instructional Technology and Distance Learning*, 12(1), 29–42. https://itdl.org/Journal/Jan_15/Jan15.pdf
- Attali, Y., & Burstein, J. (2004). Automated essay scoring with e-rater V.2.0. Educational Testing Service. https://www.ets.org/Media/Products/e-rater/erater_IAEA.pdf
- Balaji, M. S., & Chakrabarti, D. (2010). Student interactions in online discussion forum: Empirical research from “media richness theory” perspective. *Journal of Interactive Online Learning*, 9(1), 1–22. <https://www.ncolr.org/jiol/issues/pdf/9.1.1.pdf>
- Bataineh, R. F., & Mayyas, M. B. (2017). The utility of blended learning in EFL reading and grammar: A case for Moodle. *Teaching English with Technology*, 17(3), 35–49. <http://files.eric.ed.gov/fulltext/EJ1149423.pdf>
- Bateson, G. (2019). Essay (auto-grade) question type. [https://docs.moodle.org/38/en/Essay_\(auto-grade\)_question_type](https://docs.moodle.org/38/en/Essay_(auto-grade)_question_type)
- Borich, G. D. (2013). *Effective teaching methods: Research-based practice* (8th ed.). Pearson.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Longman.
- Cebrián Robles, D., Serrano Angulo, J., & Cebrián De La Serna, M. (2014). Federated eRubric service to facilitate self-regulated learning in the European university model. *European Educational Research Journal*, 13(5), 575–583. <https://doi.org/10.2304/eeerj.2014.13.5.575>
- Chapman, J. W., & Tunmer, W. E. (2003). Reading difficulties, reading-related self-perceptions, and strategies for overcoming negative self-beliefs. *Reading and Writing Quarterly*, 19(1), 5–24. <https://doi.org/10.1080/10573560308205>
- Cohen, D., & Sasson, I. (2016). Online quizzes in a virtual learning environment as a tool for formative assessment. *Journal of Technology and Science Education*, 6(3), 188–208. <http://dx.doi.org/10.3926/jotse.217>
- Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42(2), 214–257. <https://doi.org/10.1598/rrq.42.2.2>
- Cook, B. R., & Babon, A. (2017). Active learning through online quizzes: Better learning and less (busy) work. *Journal of Geography in Higher Education*, 41(1), 24–38. <https://doi.org/10.1080/03098265.2016.1185772>
- Dalelio, C. (2013). Student participation in online discussion boards in a higher education setting. *International Journal on E-Learning: Corporate, Government, Healthcare, and Higher Education*, 12(3), 249–271. <https://www.learntechlib.org/primary/p/37555>
- Davies, J., & Graff, M. (2005). Performance in e-learning: Online participation and student grades. *British Journal of Educational Technology*, 36(4), 657–663. <https://doi.org/10.1111/j.1467-8535.2005.00542.x>
- Dennen, V. P. (2005). From message posting to learning dialogues: Factors affecting learner participation in asynchronous discussion. *Distance Education*, 26(1), 127–148. <https://doi.org/10.1080/01587910500081376>
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2019). The Impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32(4), 481–509. <https://doi.org/10.1007/s10648-019-09510-3>
- Earl, L., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Government of Alberta.
- Elliott, J., & Yu, C. (2013). Learning studies in Hong Kong schools: A summary evaluation report on the ‘variation for the improvement of teaching and learning’ (VITAL) project. *Éducation Et Didactique*, 7(2), 147–163. <https://doi.org/10.4000/educationdidactique.1762>
- Espinosa Cevallos, L. F., & Soto, S. T. (2018). Pre-instructional student assessment. *MEXTESOL Journal*, 42(4), 1–9. http://www.mextesol.net/journal/index.php?page=journal&id_article=4438
- Fageeh, A. I. (2015). EFL student and faculty perceptions of and attitudes towards online testing in the medium of Blackboard: Promises and challenges. *JALT CALL Journal*, 11(1), 41–62. <https://doi.org/10.29140/jaltcall.v11n1.183>
- Falchikov, N. (2007). The place of peers in learning and assessment. In D. Boud & N. Falchikov (Eds.), *Rethinking Assessment in Higher Education: Learning for the Longer Term* (pp. 128–143). Routledge. <https://doi.org/10.4324/9780203964309>

- Fu, M., & Li, S. (2020). The effects of immediate and delayed corrective feedback on L2 development. *Studies in Second Language Acquisition*, 1–33. <https://doi.org/10.1017/s0272263120000388>
- Galizzi, M. (2010). An assessment of the impact of online quizzes and textbook resources on students' learning. *International Review of Economics Education*, 9(1), 31–43. [https://doi.org/10.1016/S1477-3880\(15\)30062-1](https://doi.org/10.1016/S1477-3880(15)30062-1)
- Gamage, S. H. P. W., Ayres, J. R., Behrend, M. B., & Smith, E. J. (2019). Optimising Moodle quizzes for online assessments. *International Journal of STEM Education*, 6(1), 1–14. <https://doi.org/10.1186/s40594-019-0181-4>
- Ganske, K., Monroe, J. K., & Strickland, D. S. (2003). Questions teachers ask about struggling readers and writers. *The Reading Teacher*, 57(2), 118–128. www.jstor.org/stable/20205331
- Griffin, P. (2017). Interpreting data to evaluate growth. In P. Griffin (Ed.), *Assessment for Teaching* (2nd ed., pp. 226–246). Cambridge University Press. <https://doi.org/10.1017/9781108116053.012>
- Griffin, P., & Care, E. (2015). The assessment and teaching of 21st century skills method. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st Century skills: Methods and approach* (pp. 3-33). Springer. https://doi.org/10.1007/978-94-017-9395-7_1
- Hamouda, A. (2012). An exploration of causes of Saudi students' reluctance to participate in the English language classroom. *International Journal of English Language Education*, 1(1). <https://doi.org/10.5296/ijelev.v1i1.2652>
- Heumann, C., Schomaker, M., & Shalabh. (2016). Hypothesis testing. In C. Heumann, M. Schomaker, & Shalabh (Eds.), *Introduction to statistics and data analysis: With exercises, solutions and applications in R* (pp. 209–247). Springer. https://doi.org/10.1007/978-3-319-46162-5_10
- Hoefl, M. E. (2012). Why university students don't read: What professors can do to increase compliance. *International Journal for the Scholarship of Teaching and Learning*, 6(2). <https://doi.org/10.20429/ijstl.2012.060212>
- Hohner, M., & Tsigaris, P. (2010). Alignment of two grading systems: A case study. *American Journal of Business Education*, 3(7), 93–102. <https://doi.org/10.19030/ajbe.v3i7.461>
- Juniati, S. R., Jabu, B., & Salija, K. (2018). Students' silence in the EFL speaking classroom. In A. Abduh, C. A. Korompot, A. A. Patak, & M. N. A. Asnur (Eds.), *The 65th TEFLIN International Conference: Sustainable teacher professional development in English language education: Where theory, practice, and policy meet* (pp. 90–94). Universitas Negeri Makassar. <https://ojs.unm.ac.id/teflin65/article/view/90-94/4224>
- Kayler, M., & Weller, K. (2007). Pedagogy, self-assessment, and online discussion groups. *Educational Technology & Society*, 10(1), 136–147. <https://drive.google.com/open?id=1sos78ToXmtzKQy1rgLFSazogp1GeWF4>
- Khonbi, Z. A., & Sadeghi, K. (2013). The effect of assessment type (self vs. peer) on Iranian university EFL students' course achievement. *Procedia - Social and Behavioral Sciences*, 70, 1552–1564. <https://doi.org/10.1016/j.sbspro.2013.01.223>
- Kim, A. S. N., Shakory, S., Azad, A., Popovic, C., & Park, L. (2020). Understanding the impact of attendance and participation on academic achievement. *Scholarship of Teaching and Learning in Psychology*, 6(4), 272–284. <https://doi.org/10.1037/stl0000151>
- Kovalik, C. L., & Dalton, D. W. (1999). The process/outcome evaluation model: A conceptual framework for assessment. *Journal of Educational Technology Systems*, 27(3), 183–194. <https://doi.org/10.2190/KLTH-KV8X-T959-PBRW>
- Khojasteh, L., Shokrpour, N., & Kafipour, R. (2015). EFL Students' perception toward class participation in general English courses. *Indian Journal of Fundamental and Applied Life Sciences*, 5(S2), 3198–3207. <http://www.cibtech.org/sp.ed/jls/2015/02/395-JLS-S2-378-SHOKRPOUR-STUDENTS.doc.pdf>
- Ramirez Leyva, E. M. (2003). The impact of the internet on the reading and information practices of a university student community: The case of UNAM. *New Review of Libraries and Lifelong Learning*, 4(1), 137–157. <https://doi.org/10.1080/1468994042000240287>
- Margolis, H., & McCabe, P. P. (2004). Resolving struggling readers' homework difficulties: A social cognitive perspective. *Reading Psychology*, 25(4), 225–260. <https://doi.org/10.1080/02702710490512064>
- McCracken, J., Cho, S., Sharif, A., Wilson, B., & Miller, J. (2012). Principled assessment strategy design for online courses and Programs. *The Electronic Journal of E-Learning*, 10(1), 107–119.
- Mellar, H., Peytcheva-Forsyth, R., Kocdar, S., Karadeniz, A., & Yovkova, B. (2018). Addressing cheating in e-assessment using student authentication and authorship checking systems: Teachers' perspectives. *International Journal for Educational Integrity*, 14(1). <https://doi.org/10.1007/s40979-018-0025-x>
- Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Association for Supervision and Curriculum Development (ASCD).
- Mustafa, F. (2015). Using corpora to design a reliable test instrument for English proficiency assessment. *Proceedings of the 62nd TEFLIN International Conference*, Denpasar, Indonesia. 344–352.
- Mustafa, F., Assyry, S. N., Bustari, A., & Nuryasmin, R. A. (2019). The role of vocabulary e-learning: Comparing the effect of reading skill training with and without vocabulary homework. *Teaching English with Technology*, 19(2), 21–43. <https://www.tewtjournal.org/?wpdmact=process&did=NTc1LmhvdGxpbnms>
- Nor, N. F. N., Razak, N. A., & Aziz, J. (2010). E-learning: Analysis of online discussion forums in promoting knowledge construction through collaborative learning. *WSEAS Transactions on Communications*, 9, 53–62. <http://www.wseas.us/e-library/transactions/communications/2010/89-351.pdf>
- Omotoso, O. J. (2013). Evaluation of grading systems of some tertiary institutions in Nigeria. *Information and Knowledge Management*, 3(2), 92–125. <https://www.iiste.org/Journals/index.php/IKM/article/view/4284/4352>
- Ozcan-Seniz, G. (2017). Best practices in assessment: A story of online course design and evaluation. *5th Annual Drexel Assessment Conference*, Philadelphia, PA., 28–35.
- Padayachee, P., Wagner-Welsh, S., & Johannes, H. (2018). Online assessment in Moodle: A framework for supporting our students. *South African Journal of Higher Education*, 32(5), 211–236. <https://doi.org/10.20853/32-5-2599>
- Pop, A., & Slev, A. M. (2012). Maximizing EFL learning through blending. *Procedia - Social and Behavioral Sciences*, 46, 5516–5519. <https://doi.org/10.1016/j.sbspro.2012.06.467>
- Privitera, G. J. (2018). *Statistics for the behavioral sciences* (3rd ed). Sage.
- Quinlan, T., Higgins, D., & Wolff, S. (2009). Evaluating the construct-coverage of the e-rater® scoring engine. *ETS Research Report Series*, 2009(1), 1-35. <https://doi.org/10.1002/j.2333-8504.2009.tb02158.x>
- Raposo-Rivas, M., & Gallego-Arrufat, M.-J. (2016). University students' perceptions of electronic rubric-based assessment. *Digital Education Review*, 30, 220–233. <https://doi.org/10.1344/der.2016.30.220-233>

- Richards, J. C., & Rodgers, T. S. (1986). *Approaches and methods in language teaching: A descriptive and analysis*. Cambridge University Press.
- Rodrigues, F., & Oliveira, P. (2014). A system for formative assessment and monitoring of students' progress. *Computers and Education*, 76, 30–41. <https://doi.org/10.1016/j.compedu.2014.03.001>
- Salas-Morera, L., Arauzo-Azofra, A., & García-Hernández, L. (2012). Analysis of online quizzes as a teaching and assessment tool. *Journal of Technology and Science Education*, 2(1), 39–45. <https://doi.org/10.3926/jotse.30>
- Savaşçı, M. (2014). Why are some students reluctant to use L2 in EFL speaking classes? An action research at tertiary level. *Procedia - Social and Behavioral Sciences*, 116, 2682–2686. <https://doi.org/10.1016/j.sbspro.2014.01.635>
- Shams, I. E. (2013). Hybrid learning and Iranian EFL learners' autonomy in vocabulary learning. *Procedia - Social and Behavioral Sciences*, 93, 1587–1592. <https://doi.org/10.1016/j.sbspro.2013.10.086>
- Sharma, R., Jain, A., Gupta, N., Garg, S., Batta, M., & Dhir, S. K. (2016). Impact of self-assessment by students on their learning. *International Journal of Applied and Basic Medical Research*, 6(3), 226–229. <https://doi.org/10.4103/2229-516x.186961>
- Siagian, B. A., & Siregar, G. N. S. (2018). Analisis penerapan kurikulum berbasis KKNI di Universitas Negeri Medan [The analysis of KKNI-based curriculum at Universitas Negeri Medan]. *Pedagogia*, 16(3), 327–342. <https://doi.org/10.17509/pdgia.v16i3.12378>
- Sieberer-Nagler, K. (2015). Effective classroom-management & positive teaching. *English Language Teaching*, 9(1), 163–172. <https://doi.org/10.5539/elt.v9n1p163>
- Swan, K., Shen, J., & Hiltz, S. R. (2019). Assessment and collaboration in online learning. *Online Learning*, 10(1), 45–62. <https://doi.org/10.24059/olj.v10i1.1770>
- Tauber, R. T. (2007). *Classroom management: Sound theory & effective practice* (4th ed.). Praeger.
- Tesfaye Abebe, D., & Deneke, T. (2015). Causes of students' limited participation in EFL classroom: Ethiopian public universities in focus. *International Journal of Educational Research and Technology*, 6(1), 74–89. <http://soeagra.com/ijert/ijertmarch2015/8.pdf>
- Tri, S., & Ratri, Y. (2015). Student factor influencing Indonesian student reading literacy based on PIRLS data 2011. *Journal of Education*, 8(1), 24–32.
- Wang, Y.-M., & Chen, D.-T. (2017). Assessing online discussions: A holistic approach. *Journal of Educational Technology Systems*, 46(2), 178–190. <https://doi.org/10.1177/0047239517717036>
- Wilén, W. W. (2004). Refuting misconceptions about classroom discussion. *The Social Studies*, 95(1), 33–39. <https://doi.org/10.3200/tsss.95.1.33-39>
- Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching*, 7(1), 3–14.
- Yulia, A., Husin, N. A., & Anuar, F. I. (2019). Channeling assessments in English language learning via interactive online platforms. *Studies in English Language and Education*, 6(2), 228–238. <https://doi.org/10.24815/siele.v6i2.14103>

Appendix

Revised Rubric for Process Assessment of Online Learning

Status on e-learning	Description	given	revised
Lesson (Material)			
Not attempted	Students have not opened the material.	0	0
Not completed	Students have opened the lesson but have not finished reading the material.	1	1.25
Completed too quickly	Students completed reading the material too quickly.	1	1.25
Not completed after >2 attempts	Students have opened the material more than twice but have not finished reading it.	2	2
Completed with expected duration	Students have finished reading the material after spending acceptable time.	3	3
Quiz (Exercise)			
Not attempted	Students have not opened the quiz.	0	0
In progress	Students have started the quiz but have not submitted it.	1	1.25
Scored at least 80	Students have submitted the quiz and obtained a score of at least 80%	3	3
Scored lower than 60	Students have submitted the quiz and obtained a score of lower than 60% (even after several attempts)	2	2.25
Scored 60 – 79	Students have submitted the quiz and obtained a score of between 60% and 79% (even after several attempts)	2.5	2.5
Completed too quickly	Students completed the quiz too quickly regardless of the score they obtained.	1.5	1.75
Discussion Forum			
Not participation	Students have not posted anything in the forum.	0	0
Participation by greeting	Students only greeted other students in the forum.	1	1.25
Asking an unnecessary question	Students ask for information which can be found in the material or information that is not important or not relevant to the course.	1	1.50
Giving an irrelevant response	Students respond to questions or statements by their classmates, but the response is irrelevant.	1	1
Giving a wrong but relevant answer	Students respond to questions, but do not answer the questions correctly.	2	2
Asking a good question	Students ask questions where the information cannot be found or is not clear in the material or in the instruction given by teachers.	2.5	2.5
Giving a correct but incomplete answer	Students respond to questions, but only answer the question partially.	2.5	2.5
Giving a correct and complete answer	Students respond to questions correctly and comprehensively.	3	3
Asking a well-responded question	Students ask such an important question that it attracts more than five responses from their classmates.	3	3
Average passing grade		2.40	2.35
Converted to score (100)		80	78